

Figure 1: Amino Acid and Nucleic Acid Sequences of *Hypocrea jecorina* Cel7A

CBH1 Homologs And Variant CBH1 Cellulases
Goedegebuur et al.
SN# Unassigned
Docket No. GC793-3
Sheet 1 of 14

1	GlnSerAlaCys ThrLeuGln SerGluThr HisProProLeu ThrTrpGln LysCysSer SerGlyGlyThr CysThrGln GlnThrGly SerValValle . CAGTGGCT GCACCTCCA ATCGGAGACT CACCGCCCTC TGACATGGCA GAAATGCTCG TCTGGGGCA CTGCACTCA ACAGACAGGC TCCGTGGCTA	
101	. IAspAlaAsn TrpArgTrp ThrHisAlaThr AsnSerSer ThrAsnCys TyrAspGlyAsn ThrTrpSer SerThrLeu CysProAspAsn GluThrCys . TCGAACGCCAA CTGGGCTGG ACTCACGGTA CGAACAGCG CAGGAACCTGC TAGCATGGCA AACTGGAG CTCGACCCCTA TGTCCTGACA ACGAGACCTG CGCGAAAGAAC TCGTGTCTGG ACGGTGGCC CGTACCGCTC ACCTACGGAG TTACCAAGG CGGTACACAGC CTCCTCATTG GCTTGTAC CCAGTCGTGC	
201	. AlaLysAsn CysCysLeuAsp GlyAlaAla TyrAlaSer ThrTyrglyVal ThrThrSer GlyAsnSer LeuSerIleGly PheValThr GlnSerAla GlnLysAsnVal GlyAlaArg LeuTyrlLeu MetAlaSerAsp ThrThrTyr GlnGluPhe ThrLeuLeuGly AsnGluPhe SerPheAsp ValAspValSer CAGAAGAAC TTGGGCTCG CCTTTACCT ATGGCAGGG ACACGACCTA CCAGGAATTC ACCCTGCTTG GCAACGAGTT CTCTTCGAT GTTGATGTT	
301	. SGlnLeuPro CysGlyLeu AsnGlyAlaLeu TyrPheVal SerMetAsp AlaAspGlyGly ValSerLys TyrProThr AsnThrAlaGly AlaLysTyr . CGCAGCTGCC GTGCGGCTTG AACGGAGCTC TCTACTTCGT GTCCATGGAC GCGGATGGTG GCGTGAGCAA GTATCCACC AACACGGCTG GCGCCAAAGTA	
401	. GlyThrGly TyrCysAspSer GlnCysPro ArgAspLeu LysPheIleAsn GlyGlnAla AsnValGlu GlyTrpGluPro SerSerAsn AsnAlaAsn CGGCAACGGGG TACTGTGACA GCCAGTGTCC CGCGGATCTG AAAGTCATCA ATGGCAGGG CAACGTTGAG GGCTGGGAGC CGTCATCCAA CAACGGAAAC ACGGGCATTTG GAGGACACGG AAGCTGTGC ATCTGAGATGG ATATCTGGGA GGCCAACCTCC ATCTCGAGG CTCCTAACCCC CCACCCCTGC ACGACTGTGC	
501	. ThrGlyIleGly GlyHiGly SerCysCys SerGluMetAsp IleTrpGlu AlaAsnSer IleSerGluAla LeuThrPro HisProCys ThrThrValGly . GGInGluIle CysGluGly AspGlyCysGly GlyThrTyr SerAspAsn ArgTyrglyGly ThrcysAsp ProAspGly CysAspTrpAsn ProTyArg . GCGAGGAGAT CTGCGAGGGT GATGGGTTGCG GCGGAACTTA CTCCGATAAAC AGATATGGCG GCACTTGCA TCCCGATGGC ACCCATACCG	
601	. LeuGlyAsn ThrSerPheTyr GlyProGly SerSerPhe ThrLeuAspThr ThrLysLys LeuThrVal ValThrGlnPhe GluThrSer GlyAlaAlle CCTGGGCAAC ACCAGCTCTC ACGGCCCTGG CTCAAGCTTT ACCCTCGATA CCACCAAGAA ATTGACCGTT GTCACCCAGT TCGAGACGTC GGGTGCCATC	
701	. AsnArgTyrTyr ValGlnAsn GlyValThr PheGlnGlnPro AsnAlaGlu LeuGlySer TyrSerGlyAsn GluLeuAsn AspAspTyr CysThrAlaGlu . AACCGATACT ATGTCAGAA TGGCGTCACT TTCCAGGAGC CCAACGCCGA GCTTGGTAGT TACTTGGCA ACCAGCTCAA CGATGATTAC TGCACAGCTG	
801	. GGLuAlaGlu PheGlyGly SerSerPhe AspLysGly GlyLeuThr GlnPhelysLys AlaThrSer GlyGlyMet ValLeuValMet SerLeuTrp . AGGAGGAGA ATTGGGGAA TCCTCTTCT CAGACAAGGG CGGCCTGACT CAGTCAAGA AGGCTACCTC TGCGGGCATG GTTCTGGTCA TGAGTCGTG	
901	. AspAspTyr TyrAlaAsnMet LeuTrpLeu AspSerThr TyrProThrAsn GluThrSer SerThrPro GlyAlaValArg GlySerCys SerThrSer GGATGATTAC TAGGCCAACAA TGCTGTGGCT GGACTCCACC TACCCGACAA ACGAGACCTC CTCCACACCC GGGAAAGCTG CTCCACCCAGC	
1001	SerGlyValPro AlaGlnVal GluSerGln SerProAsnAla LysValThr PheSerAsn IleLysPheGly ProlleGly SerThrGly AsnProSerGly . TCCGGGTGCC CTGCTCAGGT CGAATCTCACT CCAAGGGTCAC ATCAAGTTCG GACCCATTTGG CAGCACCGGC AACCCCTAGCG	
1101	. GGLYAsnPro ProGlyGly AsnProProGly ThrArgArg ProAlaThrThr ThrglySer SerProGly ProThrGlnSer HistTyrgly . GCGGCAACCC TCCGGGGAA AACCCGCTG GAACCAACC CCAGGCCACTA CACCGCCGC CACCCGGCA CCTACCCAGT CTCACCTACGG	
1201		
1301		
1401	. GlnCysGly GlyIleGlyTyr SerGlyPro ThrValcys Alaserglylhr ThrcysGln ValLeuAsn ProTyrrSer GlnCysLeu CCAGTGGGGC GGATGGCT ACAGGGCCCC CACGGCTGC GCACTTGCAAC CACTACTACT CTCAGTGCCT G	

Figure 2A: Multiple alignment of the CBH1 homologous sequences.

		1		
T. reesei mat	(1)	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC		
H. orientalis mat	(1)	QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC		
H. schweinitzii mat	(1)	QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC		
T. koninlangbra mat	(1)	QSACTIQAETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSTTNC		
T. pseudokoningii mat	(1)	QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC		
Consensus	(1)	QSACTLQTEHPSLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNC		
		50		
T. reesei mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNLSIGFVTQSA		
H. orientalis mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA		
H. schweinitzii mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA		
T. koninlangbra mat	(51)	YDGNTWSSTLCPDNECAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSQ		
T. pseudokoningii mat	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA		
Consensus	(51)	YDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSADSLSIGFVTQSA		
		100		
T. reesei mat	(101)	QKNGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD		
H. orientalis mat	(101)	QKNGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD		
H. schweinitzii mat	(101)	QKNGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD		
T. koninlangbra mat	(101)	QKNGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD		
T. pseudokoningii mat	(101)	QKNGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD		
Consensus	(101)	QKNGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVSMD		
		150		
T. reesei mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN		
H. orientalis mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN		
H. schweinitzii mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN		
T. koninlangbra mat	(151)	ADGGVSKYPSNTAGAKYGTGYCDSQCPRDLKFINGEANVEGWEPESSNNAN		
T. pseudokoningii mat	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGEANVEGWEPESSNNAN		
Consensus	(151)	ADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPESSNNAN		
		200		
T. reesei mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSND		
H. orientalis mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDGCGGTYSND		
H. schweinitzii mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDGCGGTYSND		
T. koninlangbra mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQAI CDGDGC GGTY SDD		
T. pseudokoningii mat	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDS CGGTYSGD		
Consensus	(201)	TGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICDGDGCGGTYS D		
		250		
T. reesei mat	(251)	RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVQTQFETSGAI		
H. orientalis mat	(251)	RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVQTQFETSGAI		
H. schweinitzii mat	(251)	RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVQTQFETSGAI		
T. koninlangbra mat	(251)	RYGGTCDPDGCDWNPYRLGNTSXYPGPSSFTLDTTKLMVVQTQFATSGAI		
T. pseudokoningii mat	(251)	RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVQTQFETSGAI		
Consensus	(251)	RYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVQTQFETSGAI		
		300		
T. reesei mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLT		
H. orientalis mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLT		
H. schweinitzii mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLT		
T. koninlangbra mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLT		
T. pseudokoningii mat	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLT		
Consensus	(301)	NRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLT		
		350		

Figure 2B: Multiple alignment of the CBH1 homologous sequences.

	351	400
T. reesei mat	(351) QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
H. orientalis mat	(351) QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
H. schweinitzii mat	(351) QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
T. koninlangbra mat	(351) QFKQATSGGMVLVMSLWDDYYANMLWLDSIYPTNETSSTPGAARGSCSTS	
T. pseudokoningii mat	(351) QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
Consensus	(351) QFKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTS	
	401	450
T. reesei mat	(401) SGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSGGNPPGG-NPPGTTTR	
H. orientalis mat	(401) SGVPAQLESQSPNAKVYSNIKFGPIGSTGNPSGGNPPGG-NPPGTTTR	
H. schweinitzii mat	(401) SGVPAQLESQSPNAKVVYSNIKFGPIGSTGNPSGGNPPGG-NPPGTTTR	
T. koninlangbra mat	(401) SGVPAQLESQSTNAKVVFSNIKFGPIGSTGNSSGGNPPGGGNPPGTTTR	
T. pseudokoningii mat	(401) SGVPAQLESQSSNAKVVYSNIKFGPIGSTGNSSGGSPGGGNPPGTTTR	
Consensus	(401) SGVPAQLESQS NAKVVYSNIKFGPIGSTGNPSGGNPPGG NPPGTTTR	
	451	499
T. reesei mat	(450) RPATTTGSSPGPTQSHYGQCIGGIGYSGPTVCASGTTQ-VLNPyYSQCL	
H. orientalis mat	(450) RPATTTGSSPGPTQTHYGQCIGGIGYSGPTVCASGTTQ-VLNPyYSQCL	
H. schweinitzii mat	(450) RPATTTGSSPGPTQTHYGQCIGGIGYSGPTICASGTTCQQVLNEYYSQCL	
T. koninlangbra mat	(451) RPATTTGSSPGPTQTHYGQCIGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
T. pseudokoningii mat	(451) RPATSTGSSPGPTQTHYGQCIGGIGYSGPTVCASGSTCQ-VLNEYYSQCL	
Consensus	(451) RPATTTGSSPGPTQTHYGQCIGGIGYSGPTVCASGTTQ VLNEYYSQCL	

CGTCATCTCG	GCCTTCTTGG	CCACGGCCCC	TGCTCAGTCG	GCCTGCACTC	50
TCCAAACGGA	GACTCACCCG	TCTCTGACAT	GGCAGAAATG	CTCGTCTGGC	100
GGCACTTGCA	CCCAGCAGAC	AGGCTCCGTG	GTCATCGACG	CCAACTGGCG	150
CTGGACTCAC	GCGACTAAC	GCAGCACGAA	CTGCTACGAC	GGCAACACTT	200
GGAGCTCAAC	CCTATGCCCT	GACAACGAGA	CTTGCACGAA	GAATTGCTGC	250
CTGGACGGTG	CCGCCTATGC	GTCCACGTAC	GGAGTCACCA	CGAGTGCCGA	300
CAGCCTCTCC	ATCGGCTTCG	TCACGCAATC	TGCACAGAAG	AACGTTGGCG	350
CCCGTCTCTA	CCTGATGGCG	AGTACACGA	CTTACCAGGA	GTTCACGCTG	400
CTTGGCAACG	AGTTCTCTT	TGACGTTGAT	GTTTCGCAGC	TGCCGTAAGT	450
GACAACCATT	CCCCGCGAGG	CCATCTCTC	ATTGGTTCCG	AGCTGACCCG	500
CCGATCTAAG	ATGTGGCTTG	AACGGCGCTC	TGTACTTCGT	GTCTATGGAT	550
GC GGATGGTG	GC GTGAGCAA	GT ATCCCACC	AA CACCGCCG	GC GCCAAGTA	600
CGGCACGGGC	TACTGCGACA	GCCAGTGCCC	CCGCGATCTC	AAGTTCATCA	650
ACGGCCAGGC	CAACGTTGAA	GGCTGGGAGC	CGTCCTCCAA	CAACGCCAAC	700
ACGGGTATTG	CGGGACACGG	AAGCTGCTGC	TCTGAGATGG	ATATCTGGGA	750
GGCCAACCTCC	ATCTCCGAGG	CTCTGACTCC	TCACCCCTGC	ACGACTGTTG	800
GCCAGGAGAT	CTGCGACGGT	GACGGCTGCG	GC GGAA CCTA	CTCCAACGAC	850
CGATATGGTG	GTACTTGC	GA TCCTGATGGT	TGTGATTGGA	ATCCATACCG	900
CTTGGGCAAC	ACCAGCTTCT	ATGGCCCTGG	CTCGAGCTTC	ACCCTCGATA	950
CCACCAAGAA	GTTGACCGTT	GTCACCCAGT	TCGAGACCTC	GGGTGCCATC	1000
AACCGTTACT	ATGTCCAGAA	CGGCGTCACT	TACCA GCAAC	CCAACGCCGA	1050
GCTCGGTAGT	TACTCTGGTA	ATGAGCTCAA	CGATGACTAC	TGCACAGCTG	1100
AGGAGTCGGA	ATT CGGCGGC	TCCTCCTTCT	CGGACAAGGG	CGGCCTTACT	1150
CAGTTCAAGA	AGGCCACTTC	CGGCGGCATG	GTCCTGGTCA	TGAGCTTGTG	1200
GGATGACGTG	AGTTGATAGA	CAGCATTAC	ATTGTCGTTG	GAAAGACGGG	1250
CGGCTAACCG	AGACATATGA	TATCTAACAG	TACTACGCCA	ACATGCTGTG	1300
GCTGGACTCC	ACCTACCGA	CAAACGAGAC	CTCCCTCCACC	CCC GGCGCCG	1350
TGCGCGGAAG	CTGCTCCACC	AGCTCCGGCG	TCCCCGCTCA	GCTCGAGTCC	1400
CAGTCCCCCA	ACGCCAAGGT	CGTCTACTCC	AA CATCAAGT	TCGGGCCCAT	1450
TGGCAGCACC	GGCAACCCCA	GGCGCGGAAA	CCCTCCTGGC	GGAAACCCCTC	1500
CCGGCACCCAC	CACCACCCGC	CGCCCAGCTA	CCACCACTGG	AAGCTCTCCC	1550
GGACCTACTC	AGACTCACTA	CGGCCAGTGC	GGCGGCATCG	GCTACAGCGG	1600
CCCTACGGTC	TGCGCCAGCG	GCACGACCTG	CCAGG		1635

Figure 3: *H. oreintalis* genomic DNA sequence.

Figure 4A: *H. orientalis* amino acid signal sequence.

QSACTLQQTET HPSLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC	50
YDGNTWSSTL CPDNETCAKN CCLDGAAVAS TYGVVTSADS LSIGFVTQSA	100
QKNVGARLYL MASDTTYQEF TLLGNEFSFD VDVSQLPCGL NGALYFVSMD	150
ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KFINGQANVE GWEPPSSNNAN	200
TGIGGGHGSSC SEMDIWEANS ISEALTPHPC TTVGQEICDG DGCGGTYSND	250
RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKKLTV VTQFETSGAI	300
NRYVVQNGVT YQQPNAELGS YSGNELNDDY CTAEESEFGG SSFSDKGGLT	350
QFKKATSGGM VLVMMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS	400
SGVPAQLESQ SPNAKVVYSN IKFGPIGSTG NPSGGNPPGG NPPGTTTRR	450
PATTTGSSPG PTQTHYGQCG GIGYSGPTVC ASGTTTCQVLN PYYSQCL	497

Figure 4B: *H. orientalis* mature amino acid sequence.

TCGGCCTGCA	CTCTCCAAAC	GGAGACTCAC	CCGTCTCTGA	CATGGCAGAA	50
ATGCTCGTCT	GGCGGCACCT	GCACCCAGCA	GACAGGCTCC	GTGGTCATCG	100
ACGCCAAGTG	GCGCTGGACT	CACGCTACTA	ACAGCAGCAC	GAACTGCTAC	150
GACGGCAACA	CTTGGAGCTC	AACCCCTGTGC	CCTGACAATG	AGACTTGCGC	200
GAAGAACTGC	TGCCTGGACG	GTGCCGCTA	TGCGTCCACG	TACGGAGTC	250
CCACCGAGTGC	CGACAGCCTC	TCCATCGGCT	TCGTGACACA	GTCTGCACAG	300
AAAAACGTTG	GCGCCCGTCT	CTACCTGATG	GCGAGTGCAC	CGACTTACCA	350
GGAGTTCACG	CTGCTTGGCA	ACGAGTTCTC	ATTGACGTT	GATGTTTCGC	400
AGCTGCCGTA	<u>AGTGACAACC</u>	<u>ATTCCCCCGA</u>	<u>CGCCATCTTC</u>	<u>TCATTGGTTC</u>	450
<u>GAAGCTGACC</u>	<u>CGCCGATCTA</u>	<u>AGATGTGGCT</u>	<u>TGAACGGCGC</u>	<u>TCTTTACTTC</u>	500
GTGTCCATGG	ACGCAGATGG	TGGCGTGAGC	AAGTATCCC	CCAACACCGC	550
CGGGCGCAAG	TACGGCACGG	GCTACTGTGA	CAGCCAGTGC	CCCCGCGATC	600
TCAAGTTTAT	CAACGGCCAG	GCCAACGTTG	AAGGCTGGGA	GCCGTCCTCC	650
AACAAACGCCA	ACACGGGTAT	TGGCGGACAC	GGAAGCTGCT	GCTCCGAGAT	700
GGATATCTGG	GAGGCCAACT	CCATCTCCGA	GGCTCTTA	CCTCACCCCT	750
GCACGAATGT	TGGCCAGGAG	ATCTGCGACG	GTGACGGCTG	CGGGCGAAC	800
TACTCCAACG	ACCGATATGG	TGGTACTTGC	GATCCTGATG	GTTGTGATTG	850
GAATCCATAC	CGCTTGGCA	ACACCAGCTT	CTATGGCCCT	GGCTCGAGCT	900
TCACCCCTCGA	TACCACCAAG	AAGTTGACCG	TCGTCACCC	GTTCGAGACT	950
TCGGGTGCCA	TCAACCGTTA	CTATGTCCAG	AATGGCGTC	CTTACCAAGCA	1000
ACCCAACGCC	GAGCTCGGCA	GTTACTCTGG	TAATGAGCTC	AACGATGCC	1050
ACTGCACAGC	TGAAGAGTCG	GAATTGGCG	GTTCCCTCCT	CTCGGACAAG	1100
GGCGGCCCTTA	CTCAGTTCAA	GAAGGCCACT	TCCGGCGGC	TGGTCCCTGGT	1150
CATGAGCTTG	<u>TGGGATGACG</u>	<u>TGAGTCCATA</u>	<u>GAACAGCATT</u>	<u>CACATTGTCG</u>	1200
<u>TCGGAAAGAC</u>	<u>GGCGGCTAA</u>	<u>CCGAGACATT</u>	<u>ACAGTACTAC</u>	<u>GCCAACATGC</u>	1250
TGTGGCTGGA	CTCCACCTAC	CCGACAAACG	AGACCTCCTC	CACCCCCGGT	1300
GCCGTGCGCG	GAAGCTGCTC	CACCAGCTCC	GGCGTCCCAG	CTCAGCTCGA	1350
GTCCCAGTCC	GCCAACGCCA	AGGTCGTCTA	CTCCAACATC	AAGTTCGGAC	1400
CCATTGGCAG	CACCGGCAAC	CCCAGCGGG	GAAACCCCTCC	TGGCGGAAAC	1450
CCTCCCGGCA	CCACCACCA	CCGCCGCCCA	GCTACCACCA	CTGGAAGCTC	1500
TCCCGGACCT	ACTCAGACTC	ACTATGGCCA	GTGCGGCGGC	ATCGGCTACA	1550
GCGCCCTAC	GATCTGCGCC	AGCGGCACCGA	CCTGCCAGG		1589

Figure 5: *H. scweinitzii* genomic DNA sequence.

Figure 6A: *H. Schweinitzii* signal peptide.

QSACTLQTET HPSLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC	50
YDGNTWSSTL CPDNETCAKN CCLDGAAAYAS TYGVVTSADS LSIGFVTQSA	100
QKNVGARLYL MASDTTYQEF TLLGNEFSFD VDVSQQLPCGL NGALYFVSM	150
ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KFINGQANVE GWEPSNNAN	200
TGIGGHGSSC SEMDIWEANS ISEALTPHPC TNVGQEICDG DGCGGTYSND	250
RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTKKLTV VTQFETSGAI	300
NRYVYQNGVT YQQPNAELGS YSGNELNDAY CTAEESEFGG SSFSDKGGLT	350
QFKKATSGGM VLVMMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS	400
SGVPAQLESQ SANAKVVVYSN IKFGPIGSTG NPSGGNPPGG NPPGTTTTRR	450
PATTTGSSPG PTQTHYGQCG GIGYSGPTIC ASGTTCQVLN PYYSQCL	497

Figure 6B: *H. Schweinitzii* mature amino acid sequence. 497 residues

TCGGCCTGCA	CCATTCAAGC	GGAGACTCAC	CCGCCTCTGA	CATGGCAGAA	50
ATGCTCATCC	GGTGGTAGTT	GCACCTCGCA	AACCGGTTCT	GTGGTGATTG	100
ACCGGAACTG	GCGATGGACT	CACCGCGACTA	ACAGCACCAC	GAACTGCTAC	150
GACGGTAACA	CTTGGGAGCTC	CAGTCTTGC	CCCGACAATG	AGAGTTGCGC	200
AAAGAACTGC	TGCCTGGACG	GTGCAGCCTA	CGCATCCACG	TACGGAGTCA	250
CCACGAGTGC	TGATAGCCTC	TCCATTGGCT	TCGTCACTCA	GTCTCAGCAG	300
AAGAATGTTG	GCGCTCGTCT	CTACCTGATG	GCAAGCGACA	CGACCTACCA	350
GGAATTTACC	CTGCTTGGCA	ACGAGTTCTC	TTTCGATGTT	GATGTTTCAC	400
AGCTGCC <u>GT</u> A	<u>GT</u> GACTAGC	ATTTACCTCC	GACGCCATCT	CATTGATTCC	450
<u>CAG</u> CTGACGG	<u>CCA</u> ATTCAAG	ATGTGGCTTG	AACGGAGCCC	TTTACTTCGT	500
GTCCATGGAC	GCGGATGGTG	GCCTGAGCAA	GTATCCCTCC	AAACACTGCCG	550
GCGCCAAGTA	CGGCACGGGC	TACTGCGATA	GCCAGTGTCC	CCGTGATTG	600
AAGTTCATCA	ACGGCGAGGC	CAACGTTGAG	GGCTGGGAGC	CGGCTTCGAA	650
CAACGCCAAC	ACGGGTATTG	GCGGACACGG	AAGCTGCTGC	TCTGAGATGG	700
ATATCTGGGA	GGCCA ACT CC	ATCTCTGAGG	CCCTTACTCC	TCACCC TT GC	750
ACGACTGT CG	GCCAGGCCAT	TTGCGATGGT	GACGGCTGCG	GTGGAACCTA	800
CTCCGATGAC	CGATATGGTG	GTACTTGC G A	TCCTGATGCC	TGTGACTGGA	850
ACCCATACCG	CTTGGGCAAC	ACCAGCTTCT	ACGGCCCCGG	CTCGAGCTTC	900
ACCC CT CGACA	CCACCAAGAA	GATGACCGTC	GTCACCCAGT	TCGCTACTTC	950
GGGTGCCATC	AACCGATACT	ATGTCCAGAA	TGGCGTCACT	TTCCAGCAGC	1000
CCAACGCCGA	GCTCGGTAGC	TACTCTGGCA	ACACGCTCAA	CGATGCTTAC	1050
TGCGCAGCTG	AAGAGGCGGA	ATT CG CGGA	TCAT TT TCT	CAGACAAGGG	1100
TGGCC TT TACC	CAATTCAAGC	AGGCTACTTC	AGGCGGCATG	GTCTTGGTTA	1150
TGAGCCTGTG	GGATGAC <u>GT</u> G	<u>AG</u> TT CA TGGA	TAGCATTGAC	ATTGTCGAGA	1200
<u>GA</u> ACCATAGC	<u>CG</u> CTGACCGA	<u>GAC</u> ACAACAG	TACTACGCCA	ACATGCTGTG	1250
GCTGGACTCC	ATCTACCCGA	CGAACGAGAC	CTCCTCTACC	CCCGGTGCCG	1300
CGCGCGGAAG	CTGCTCTACC	AGCTCCGGTG	TCCCTGCCA	GCTCGAGTCT	1350
CAGTCTACCA	ACGCCAAGGT	CGTCTTCTCC	AACATCAAGT	TCGGACCCAT	1400
TGGCAGCACT	GGTAACTCCA	GCGGCGGAAA	CCCCCCC GG GC	GGAGGAAACC	1450
CCCCCGGCAC	CACCACCA	CGACGCCAG	CTACCA CC AC	CGGAAGCTCT	1500
CCCCGACCTA	CTCAGACACA	CTATGGCCAG	TGCGGTGGAA	TTGGGTACTC	1550
GGGCCCCACG	GTCTGCGCCA	GCGGCAGCAC	ATGCCAGG		1588

Figure 7: *T. konilangbra* genomic DNA.

Figure 8A: *T. konilangbra* signal sequence.

QSACTIQAET HPPLTWQKCS SGGSCTSQTG SVVIDANWRW THATNSTTNC	50
YDGNTWSSL CPDNESCAKN CCLDGAAAYAS TYGVVTSADS LSIGFVTQSQ	100
QKNVGARLYL MASDTTYQEF TLLGNEFSFD DVSQLPCGL NGALYFVSMD	150
ADGGVSKYPS NTAGAKYGTG YCDSQCPRDL KFINGEANVE GWEPAASNAN	200
TGIGGHGSCC SEMDIWEANS ISEALTPHPC TTVGQAICDG DCGGGTYSDD	250
RYGGTCDPDG CDWNPYRLGN TSXYGPGSSF TLDTKKMTV VTQFATSGAI	300
NRYYVQNGVT FQQPNAELGS YSGNTLN DAY CAAEEAEFGG SSFSDKGGLT	350
QFKQATSGGM VLVMSLWDDY YANMLWLDSI YPTNETSSTP GAARGSCSTS	400
SGVPAQLESQ STNAKVVFSN IKFGPIGSTG NSSGGNPPGG GNPPGTTTR	450
RPATTTGSSP GPTQTHYGQC GGIGYSGPTV CASGSTCQVL NPYYSQCL	498

Figure 8B: *T. konilangbra* mature amino acid sequence.

CBH1 Homologs And Variant CBH1 Cellulases

Goedgebuur et al.

SN# Unassigned

Docket No. GC793-3

Sheet 10 of 14

TCGGCCTGCA	CCCTCCAGAC	GGAAACTCAC	CCGCCTCTGA	CATGGCAGAA	50
ATGCTCATCT	GGTGGCACTT	GCACCCAACA	GACGGGCTCC	GTGGTCATCG	100
ACCGAAGTG	GCGCTGGACT	CACGCTACGA	ACAGCAGCAC	GAACTGCTAC	150
GACGGTAACA	CTTGGAGCTC	AACCTTGTC	CCTGACAATG	AGACTTGCGC	200
GAAGAACTGC	TGCTGGATG	GTGCCGCCTA	CGCGTCGACG	TACGGAGTCA	250
CCACGAGCGC	TGACAGCCTC	TCCATTGGCT	TCGTCACTCA	GTCTGCCAG	300
AAGAATGTCG	GCGCCCGTCT	CTACTTGATG	GCGAGTGACA	CGACCTACCA	350
AGAATTTACC	CTGCTTGGCA	ACGAGTTCTC	CTTCGATGTT	GATGTTCCC	400
AGCTGCCGTA	AGTGGCCAAC	TACACCCCTT	GACGGTATCC	TCTCATTGGT	450
TCCCCAGCTGA	CTCGCGAAAT	TAAGATGTGG	CTTGAACCGA	GCTCTTACT	500
TCGTGTCCAT	GGACGCGGAT	GGTGGCGTGA	GCAAGTATCC	CACAAACACT	550
GCCGGCGCCA	AGTACGGCAC	GGGTTACTGT	GACAGCCAGT	GCCCTCGTGA	600
TCTCAAGTTC	ATCAACGGCG	AGGCCAACGT	TGAGGGCTGG	GAGCCGTTCT	650
CCAACAAACGC	CAACACGGGC	ATTGGCGGAC	ATGGAAGCTG	CTGCTCTGAG	700
ATGGATATCT	GGGAGGCCAA	CTCCATCTCT	GAGGCTCTTA	CTCCTCATCC	750
TTGCACGACC	GTCGGGCAGG	AAATTGCGA	TGGTGACTCC	TGCGGGCGGAA	800
CCTACTCCGG	TGATCGATAT	GCGGGTACTT	GCGATCCTGA	TGGCTGCGAT	850
TGGAACCCAT	ACCGCTTGGG	CAACACCCAGC	TTCTACGGGC	CCGGCTCAAG	900
CTTCGCTCTT	GATAACCACCA	AGAAGTTGAC	CGTTGTCACC	CAGTTGAGA	950
CTTCGGCGC	TATCAACCGG	TACTACGTCC	AGAATGGCGT	CACTTTCCAG	1000
CAGCCCAACG	CCGAGCTCGG	TAGTTACTCT	GGCAACTCGC	TCGACGATGA	1050
CTACTGCGCG	GCTGAAGAGG	CGGAGTTGG	TGGCTCTTCT	TTCTCGGACA	1100
AGGGCGGCCT	TACTCAATT	AAAAGGCTA	CTTCGGTGG	CATGGTCTTG	1150
GTCATGAGCT	TGTGGATGA	TGTGAGTTCA	TGAATAGCAT	TCAAACAGTC	1200
AACAGAATAA	CAGCAGCTGA	CTGAGACACA	ATAGTACTAC	GCCAACATGC	1250
TGTGGCTGGA	CTCCACCTAC	CCGACGAACG	AGACCTCTTC	CACCCCCGGT	1300
GCCGTGCGCG	GAAGCTGCTC	CACCAGCTCC	GGTGTCCCTG	CTCAGCTTGA	1350
GTCCCAGTCT	TCCAACGCCA	AGGTCGTATA	CTCCAACATC	AAGTTCGGCC	1400
CTATCGGCAG	CACCGGCAAC	TCCAGCGGGC	GTAGCCCTCC	CGGCAGGAGA	1450
AACCCCTCCCG	GTACCACGAC	CACCCGCCGC	CCAGCTACCT	CCACTGGAAAG	1500
CTCTCCCGGC	CCTACTCAGA	CGCACTATGG	CCAGTGCAGT	GGTATTGGGT	1550
ACTCGGGCCC	CACGGTCTGC	GCGAGTGGCA	GCACCTGCCA	GG	1592S

Figure 9: *T. pseudokonigii* genomic DNA sequence.

Figure 10A: *T. pseudokoningii* signal sequence.

QSACTLQETP HPPLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC	50
YDGNTWSSTL CPDNETCAKN CCLDGAAYAS TYGVTSADS LSIGFVTQSA	100
QKNVGARLYL MASDTTYQEF TLLGNEFSFD VDVSQQLPCGL NGALYFVSM	150
ADGGVSKYPT NTAGAKYGTG YCDSQCPRLD KFINGEANVE GWEPEFSNNAN	200
TGIGGHGSCC SEMDIWEANS ISEALTPHPC TTVGQEICDG DSCGGTYSGD	250
RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF ALDTTKKLTW VTQFETSGAI	300
NRYYVQNGVT FQQPNELGS YSGNSLDDDY CAAEEAEFGG SSFSDKGGLT	350
QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS	400
SGVPAQLESQ SSNAKVYVSN IKFGPIGSTG NSSGGSPPGG GNPPGTTTR	450
RPATSTGSSP GPTQTHYGQC GGIGYSGPTV CASGSTCQVL NPYYSQCL	498

Figure 10B: *T. pseudokoningii* mature amino acid sequence.

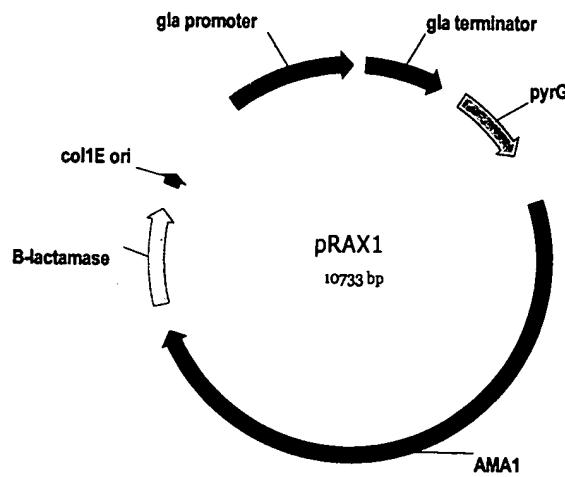


Figure 11: pRAX1

Figure 12: Destination vector pRAXdes2 for expression in *A. niger*

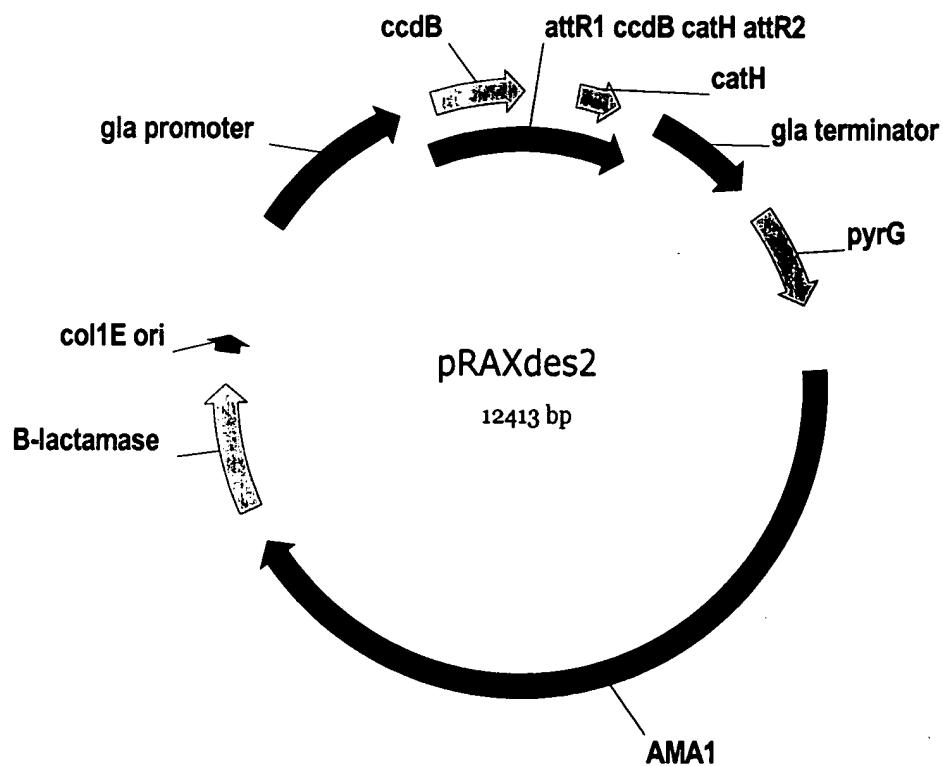


Figure 13: Replicative expression pRAXdesCBH1 vector of CBH1 genes under the control of the glucoamylase promotor.

